5P-04

インタラクティブなシーケンシャルパターンマイニングの性能向上に向けた検討

青柳結衣† Le Hieu Hanh† 松尾亮輔† 山崎友義† 荒木賢二† 横田治夫‡ 小口正人†

†お茶の水女子大学 ‡城西大学

1 はじめに

様々なビッグデータの蓄積に伴い，シーケンスを対象とする SPM (Sequential pattern mining) が注目されている．SPM とは，閾値 (minsup) を指定し，頻度の高いシーケンシャルパターンを抽出する解析手法である．シーケンスがデータセット内に出現する頻度をサポート値とし，サポート値が minsup より大きい場合に頻度が高いとみなされる．最適な minsup はデータセットに依存するため，minsup を調整しながら解析を行うのに長けているインタラクティブな SPM が不可欠である．

インタラクティブな SPM の有名なアルゴリズムとして KISP[1] が挙げられる．サポート値が既知のシーケンスを知識ベース (KB) に保存し，適宜参照することで minsup の変更に柔軟に対応できる．KISP は minsup が単調減少する場合において，高速であることが明らかになっている．しかし，単調増加する場合や増減を繰り返す場合は実行時間が増加する可能性がある．本研究では，KB 構造の再設計を行い，より効率的で高速なインタラクティブな SPM 手法を検討する．

2 関連研究

2.1 KISP

KISP は，知識ベース KB を使用するインタラクティブな SPM で，minsup を変更し繰り返し実行するのに適している．指定された minsup が KB.base より大きい場合は，minsup を満たすパターンを KB から単純に取得するため，パフォーマンスが大幅に向上する．

指定された minsup が KB.base 以上である場合は，頻出シーケンシャルパターンを KB から取得する．minsup を満たす頻出パターンは全て KB に保存されているため，データセットにアクセスせず，頻出パターンを得ることができる．

minsup が KB.base より小さい場合は新たにパターンを探索する．探索をする際，最初に新しい候補シーケンスを生成する．KISP は既存のインタラクティブな SPM より候補シーケンス数が少ないため，高速である．最後に候補シーケンスのサポート値を算出し，KB に保存する．全ての候補シーケンスについて算出後，minsup の条件を満たす頻出シーケンシャルパターンを取得できる．

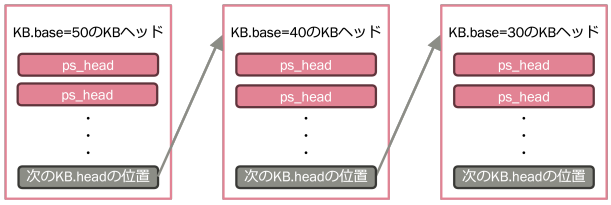

図 1: KISP の KB ヘッド例

2.1.1 KISP 内の KB 構造

KISP の KB は最小の KB.base と複数の KB ヘッドで構成されている．新しい情報が追加されるたびに KB ヘッドを作成する．KB ヘッドは，KB ヘッドが生成された時の minsup，パターンサポートテーブルを要約したパターンサポートヘッドとその総数，次の KB ヘッドへのリンクの 4 つの要素で構成されている．パターンサポートテーブルは同じサイズのパターンをグループ化して保存しており，特定のサイズのパターンを迅速に見つけることが可能である．候補シーケンス生成時にパターンの長さが重要であるため，グループ化して保存している．

KB ヘッドの構造の例を図 1に示す．最初に minsup=50 を与えると KB.base=50 の KB ヘッドが生成される．次に minsup=40 を与えると KB.base=40 の KB ヘッドを生成する．このように，KB に存在しない新しいパターンのサポート値を計算するたびに KB ヘッドが増加する．KB.base=40 の KB ヘッドには minsup=50 のシーケンスを求めるときに必要な情報と minsup=40 を求めるときに必要な情報の差分のみが挿入されている．

2.2 インタラクティブな SPM

先行研究 [2] では，クローズド頻出シーケンシャルパターンのみを対象とすることで，冗長なパターンを削減し，インタラクティブな SPM の高速化を目指した．本研究では，知識ベースの管理および参照方法を拡張し，minsup の変動に対する計算コストの削減を目指す．

Improving the Performance of Interactive Sequential Pattern Mining

†Aoyagi Yui †Le Hieu Hanh †Matsuo Ryosuke
†Yamazaki Tomoyoshi †Araki Kenji
‡Yokota Haruo †Oguchi Masato
†Ochanomizu University
‡Josai University

Copyright ©2026 Information Processing Society of Japan.
All Rights Reserved.

表 1: 実験環境

サーバ	Dell PowerEdge R740xd
CPU	Intel Xeon Gold 5218 16 cores x 2
OS	Ubuntu 24.04.1 LTS
メモリ	64GB x 6
python	3.12.3

表 2: データセットの概要

	BMSWebView1	BMSWebView2
シーケンス数	59,601	77,512
平均要素数	2.42	4.62
サイズ (MB)	1.5	3.6
内容	EC サイトのクリックストリームデータ	

3 提案手法

提案するインタラクティブな SPM では，KISP を改良している．主な提案は，KB の構造を変更することで閾値が増加する場合にも対応できるようにすることである．KISP は minsup が増加する場合は 1 番手前にある KB ヘッドにアクセスし続ける．常に先頭の KB ヘッドから minsup 条件を満たすパターンを抽出する必要があるため，コストが余分にかかる．そこで，今までの実行における最大の minsup を KB に保存する．指定した minsup がそれよりも大きい場合は，新しい KB ヘッドを作成する．例では，minsup=100 を与えると KB.base=100 の KB ヘッドを新たに作成する．

4 実験

4.1 実験概要

本実験では KB 構造変更による実行時間の検証を目的とする．表 1に実験環境を示した．使用したデータセット [3] の詳細な内容は，表 2に示した．それぞれのデータセットからランダムにシーケンスを 5,000 個抽出し，新たにデータセットを作成した．2 種類のデータセットに対し，KISP と提案手法を適用し，minsup が単調増加する場合と増減する場合の実行時間を測定した．単調増加の場合，minsup は 10 から 150 まで 10 ずつ増加させた．増減する場合では，minsup は 10 から 210 まで 40 刻みで増加させたのち，20 刻みで減少させた．

4.2 実験結果

minsup が単調増加する場合の実行時間の結果を図 2に，minsup が増減する場合の実行時間の結果を図 3に示す．

単調増加の場合では，BMSWebView1 に適用した場合は有効性があるものの，僅かな時間差となっている．よって，実行時間が減少するケースもあるが，減少幅が小さく，有効性が弱いことが確認された．提案手法は KISP より多くの KB ヘッドを作成するため，KB 構築時にコストがかかるが，KB ヘッドからシーケンスを抽出しやすいため，KISP とほぼ同等の結果を得ることができている．

実行時間	KISP(s)	提案手法(s)	提案手法/KISP
BMSWebView1	127.66	126.51	99.10%
BMSWebView2	1531.20	1535.00	100.25%

図 2: 単調増加時の実行時間比較

実行時間	KISP(s)	提案手法(s)	提案手法/KISP
BMSWebView1	129.02	126.03	97.68%
BMSWebView2	1525.86	1524.91	99.94%

図 3: 増減時の実行時間比較

増減する場合は，どちらのデータセットにおいても実行時間が短縮されている．KISP は KB.base が 10 の KB ヘッドから抽出しているのに対し，提案手法はより近い KB.base の KB ヘッドから抽出しているためだと考えられる．増減に関しては様々な minsup の与え方があり，今回の実行よりも KB ヘッドの作成数が増えてしまうと優勢ではない場合もあるため，検討の余地がある．

5 まとめと今後の課題

本研究では KB 構造を見直すことで minsup が増減する場合，実行時間が減少するケースもあることが明らかになった．

今後の課題としては，KB 構造を変更したことで実行時間が増加する場合と減少する場合がある原因の解明が挙げられる．また，KB ヘッド内の頻出パターン管理方法の検討も進めていきたい．

謝辞

本研究の一部は日本学術振興会科学研究費 (#24K02943) の助成からの支援によって行われた．

参考文献

[1] M. Y. Lin and S. Y. Lee. Improving the efficiency of interactive sequential pattern mining by incremental pattern discovery. In *International Conference on System Sciences*, pages 68–76, 2002.

[2] Yui Aoyagi, Hieu Hanh Le, Ryosuke Matsuo, Tomoyoshi Yamazaki, Kenji Araki, Haruo Yokota, and Masato Oguchi. Improving the efficiency of interactive sequential pattern mining by closed pattern discovery. In *Advanced Data Mining and Applications. ADMA 2025. Lecture Notes in Computer Science*, pages 248–256. 2026.

[3] Fournier-Viger P. An open-source data mining library. `http://www.philippe-fournier-viger.com/spmf/index.php?link=datasets.php`, 2024. Accessed: 2024-12-20.

Copyright ©2026 Information Processing Society of Japan.
All Rights Reserved.